

Why usage statistics cannot tell us everything, and why we shouldn't dare to ask

Final manuscript published in *Against the Grain* 15(6), 24-26, Jan 2004

Philip M. Davis
Life Sciences Librarian
Cornell University
pmd8@cornell.edu

Introduction

What would you be willing to pay for a full-text download?
\$30, \$20, \$5, 25 cents?

Would you answer this question differently if you knew:

- 1) That it was a faculty member needing the article while preparing a million-dollar research grant application?
- 2) That it was a graduate student doing a comprehensive literature search for her dissertation?
- 3) That it was an undergraduate downloading the first three documents for an assignment due tomorrow?

Would you also answer the question differently if you knew:

- 1) That you were paying for an article download that you already had in your stacks?
- 2) That this article was downloaded first as an HTML document and then as a PDF (i.e. two downloads)?
- 3) That the same article was available in a competing product for a tenth of the cost?

And lastly, would you change how much you would be willing to pay if you knew:

- 1) That the article was actually read
- 2) That only the title and abstract were browsed
- 3) That someone systematically downloaded this article (along with every other article in this journal) because he wanted to create a personal archive?

While usage statistics can tell us so much about *how much* a journal or resource is being used, it cannot tell us *why* it was used, or *by whom*. Asking the *why* and *by whom* questions may be very tempting for both librarians and publishers. I will argue in this short piece that the answer to these questions have unintended consequences for library budgets.

Price Discrimination

Price discrimination is charging different prices to different customers for the same good or service. It is practiced because it results in greater profits than if a company charged a uniform price to all customers. Price discrimination is practiced at the movie theatre (age discrimination), for a haircut (gender discrimination), and for airline tickets (by charging business travelers more than vacationers). Consumers often despise price discrimination. Those who pay more don't believe that it is "fair" to pay more for the same service, and those who pay less may feel that they could have received a better deal.

The electronic publishing market is ideal for price discrimination. Infrastructure costs are very high and marginal costs are exceedingly low. The cost to create first online copy is the publishers' greatest expense, while the cost to deliver one more PDF or HTML article to a desktop are almost so low as to consider them to be non-existent. Because the marginal costs are so small, a publisher can sell (or bundle) online access to institutions who have never owned a paper copy at prices that sound too good to be true.

Price discrimination only works when the producer has some monopoly power, which means that the customer cannot equally substitute one product or service for another. In an exact sense, each journal is a monopoly since it is composed of a collection of unique articles not found in any other product -- in essence, there is no substitutability. In a practical sense this is not completely true. Undergraduates will often require three articles on a particular topic, and so the substitutability for any article (as long as it is on the topic) is very high.

There are three different *degrees* of price discrimination, and the distinctions between these are important for the publishing economy.

First Degree

First-degree price discrimination means that a producer (publisher) can sell every unit (journal article) at the maximum price that each consumer is willing to pay. In reality, a publisher currently lacks the information necessary to put this pricing model into practice. When you try to download an article from a journal to which your library does not subscribe, you (and everyone else) are pitched the same price.

In the case of journal subscriptions, the use of subscription vendors prohibited publishers from knowing too much about their customers. Publishers however, were not entirely in the dark. The construction of the ISI Citation Reports in the mid 1970's provided publishers with a good idea of how important their journals were to the academic community as a whole. These citations reports however did not allow publishers to find out how their journals were being used at *each institution*. In effect, they did not have enough information to charge differential prices to each institution based on what publishers believed libraries were willing to pay.

Second Degree

Second-degree price discrimination is where the customer is offered a choice of possible contracts that reveal information about themselves through their choice. Booking a flight months in advance to the Charleston Conference and including a Saturday stay-over distinguishes you from most business travelers who book very close to the travel date and often fly during the week. In this case, your choice reveals information about yourself and allows the producer to charge very different prices for the same service.

Third Degree

Third-degree price discrimination is where a producer can identify different types of customers and offer different contracts to each group based on their willingness to pay. Movie theatres offer one price to adults, and a discounted price to students or seniors. Many publishers distinguish different types of subscribers. For example, the New England Journal of Medicine sets different prices depending on whether you are a physician, student, resident, institution, or other. You are also required to disclose your country, and if subscribing as an individual, required to disclose your place of work or study. Institutional subscribers to PNAS are grouped by Carnegie classification, and distinguish academic from commercial institutions. The Ecological Society of America involves a more complex matrix for institutional subscribers involving country AND institution type. Many journals publishing online with HighWire Press offer free online access to developing countries. Other pricing classifications may include FTE counts (or subsets of faculty within a discipline). The practice of third-degree price discrimination is not limited to commercial publishers, and may not be considered “unfair”. By charging different amounts to different subscribers a non-profit publisher may use commercial subscribers to help subsidize member subscriptions, student subscriptions, or to help run other society services.

Why knowing too much about our patrons is detrimental for library budgets (and really good for publishers)?

As mentioned above, academic publishers have been practicing third-degree price discrimination for quite some time, offering differential pricing for classes of users. While more customer surplus can be extracted by practicing this type of discrimination over setting a uniform price, it is not nearly efficient as first-degree price discrimination.

In order for publishers to move into first-degree price discrimination, they need to know more information about how their product is valued at each institution. Raw number of downloads would allow publishers to compare institutions and their use of the same product. Knowing *who* used their product and *how* or *why* they used it would allow greater leverage to price discriminate. At present, both librarians and publishers are only privy to anonymous download statistics. We don't know whether it was a faculty

member, a cataloger, or a student who downloaded an article, and certainly didn't know whether the download was for the purpose of research, verifying whether the resource works, or a random article pulled from a database search. If publishers knew this information, they could use it to move toward first-degree price discrimination and profit maximization. Charging libraries different amounts for downloads based on *who* and *why* they downloaded the article would yield more profits than setting a uniform cost/download for all articles.

Defending Patron Privacy Can Have Same Results

Most librarians adamantly defend the right of patron privacy, and refuse in principle to sign licenses that require patrons to log on to publisher products. I say "most" since some of my colleagues do not think this is as important an issue as some believe. The unintended consequence of preserving patron confidentiality prevents personal information being gathered and analyzed by the publisher, who could then turn this knowledge into extracting consumer surpluses from each library -- in other words, price maximize at each institution.

Other ways to gather personal information

There are other ways of gathering personal information about the user without requiring mandatory sign-in (and the subsequent wrath of librarians). One way is to provide added services that go beyond the mere content of the journal. The New England Journal of Medicine offers a host of value-added services (like continuing medical education, table of contents services, personal customization services). In order to make these features available, the publisher must be able to clearly identify an individual, require e-mail address, and may ask for additional personal information. Since the individual who uses these services does not pay the direct fees for the subscription, there is no disincentive for to withhold this information. The only caveat is that personal disclosure for services may not be accurately provided. An American publisher colleague of mine is registered for access to the NY Times online as a Chilean vintner -- an illustration that many of us may not take personal disclosure very seriously.

Conclusion

Publishers currently practice third-degree price discrimination whereby different classes of subscribers are charged different prices based on their ability (or willingness) to pay. Distinguishing the type of user within the institution and determining why a resource was used would allow publishers to start practicing first-degree price discrimination -- an economic environment where publishers can start maximizing profits from each institution. Protecting patron privacy also has unintended consequences of protecting the library budget.